



Sheikh Akbari, A., Canagarajah, CN., Redmill, DW., & Bull, D. (2008). *Disparity compensated view filtering wavelet based multiview image code using Lagrangian optimization*. 309 - 312.
<https://doi.org/10.1109/3DTV.2008.4547870>

Peer reviewed version

Link to published version (if available):
[10.1109/3DTV.2008.4547870](https://doi.org/10.1109/3DTV.2008.4547870)

[Link to publication record in Explore Bristol Research](#)
PDF-document

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

DISPARITY COMPENSATED VIEW FILTERING WAVELET BASED MULTIVIEW IMAGE CODEC USING LAGRANGIAN OPTIMIZATION

Akbar Sheikh Akbari, Nishan Canagarajah, David Redmill, and David Bull

Department of Electrical and Electronic Engineering, University of Bristol, BS8 1UB, UK

ABSTRACT

This paper presents a disparity compensated view filtering wavelet based multiview image coding scheme. The proposed codec decorrelates the input views into their frequency subbands using a disparity compensated lifting based wavelet transform. The codec then applies a 2D wavelet transform on baseband-image transforming it into a number of subbands. The general form of the Lagrangian rate distortion optimization algorithm is then modified and used along with the steepest descent algorithm to assign bits among the different subbands in such a way that it minimizes the PSNR variance of the decoded images. Two sets of experimental results are generated using three sets of multiview test images. In the first set of experiments the effect of using a weighted λ for high frequency images on the PSNR variance of the decoded images is investigated. In the second set of experiments, performance of the proposed codec in reducing the variance of the PSNR of the decoded images is investigated. Results indicate that the proposed codec gives lower PSNR variance among the decoded images compared to the basic form of the codec at all bitrates.

Index Terms— Image codecs, multidimensional coding, and stereo vision

1. INTRODUCTION

3D and free viewpoint images are new types of natural media that expand the user's sensation far beyond what is offered by traditional media. However, the price for utilizing the natural image media enormously increases the amount of data to be stored or transmitted. Since multiview images are captured by multiple cameras from a world scene and different view points, there is much correlation between the multi-view images. Therefore efficient compression schemes are vital to exploit the redundancy among the multiview images to reduce the size of the compressed dataset and realize 3D imaging systems [1]-[2]. Traditionally block based disparity compensated prediction methods are used to code the multiview images and have shown a strong capability in dealing with such amounts of data [1]. As these schemes generally use a Lagrangian rate distortion

optimization to assign the optimal rate among the subbands and the Lagrangian rate distortion algorithm minimizes the global distortion among the decoded images, the decoded images exhibit different visual quality, which is not acceptable in some applications.

In this paper, a novel disparity compensated view filtering wavelet based multiview image compression scheme is presented. The proposed codec applies a disparity compensated lifting based wavelet transform on the input images and decomposes the images into a number of high and low frequency images. The codec then performs a 2D DWT on the baseband image and transforms the baseband image into its spatial subbands. The proposed codec finally uses a modified version of the Lagrangian rate distortion optimization along with the steepest descent algorithm to assign optimal bits among the different subbands in such a way that the decoded images exhibit almost the same visual quality. The rest of this paper is organized as follows: in Section 2 the disparity compensation view filtering is discussed; Section 3 presents the proposed codec; experimental results are given in Section 4; finally Section 5 concludes the paper.

2. DISPARITY COMPENSATED VIEW FILTERING

Disparity compensation (DC) is a key technique that most of the multiview image codecs employ to exploit the correlation between the different views. These codecs employ a lifting based wavelet transform to exploit the interview redundancies between multiview images and decompose the images into their frequency subbands. The wavelet transform concentrates most of the images' energy into the baseband-image while a small portion of their energy goes to the high frequency images. Integration of DC and the wavelet transform can be well established using the lifting based wavelet transform.

The lifting based wavelet decomposition can be implemented through three basic steps called split, prediction and update. In the first step, the multiview images are split into even and odd image groups, denoted as A- and B-image groups. In the second step, A-images are predicted from the B-images using a prediction filter, $P(z)$, and then the high-pass residual images, H-images, are generated by subtracting the predicted images from the original images.

Finally, an update step combines the H-image, which has been filtered with an update filter, $U(z)$, and the corresponding B-image generates the low-pass image, L-image. The update step reduces the effect of aliasing in the L-images [3]. Figure 1 shows the implementation steps of a 5/3 lifting based wavelet transform. In this figure $P(z) = -\frac{1}{2}z_1^k z_2^l$ and $U(z) = \frac{1}{4}z_1^{\bar{k}} z_2^{\bar{l}}$ are prediction and update filters of the 5/3 wavelet transform, where k, l and \bar{k}, \bar{l} describe the DC and its Inverse Disparity Compensation (IDC) shift values, respectively. Figure 2 shows the implementation steps of a lifting based Haar wavelet transform. In this figure the prediction and update filters of the Haar wavelet transform are $P(z) = -z_1^k z_2^l$ and $U(z) = \frac{1}{2}z_1^{\bar{k}} z_2^{\bar{l}}$, where k, l and \bar{k}, \bar{l} describe the DC and IDC shift values, respectively [3].

In this paper, block based disparity estimation/compensation with one pixel resolution are used to implement the lifting based wavelet transforms. Mean Square Error (MSE) measure is also employed to choose the best matching block during disparity estimation/compensation. Disparity Vectors (DVs), which are calculated during disparity estimation/compensation, have to be inverted for IDC, which is normally impossible, as there might be areas in the reference image that may not be referred to at all and also there are areas which may be referred to more than once. However, in the proposed codec the DVs are inverted and used for IDC and in the case that a block of pixels is pointed more than once, the first time this occurs is used. Although the DC and IDC operations are not exactly inverse of each other, the lifting structure of the wavelet transform guarantees the invariability of the transform [4].

3. MULTIVIEW IMAGE ENCODER

Figure 3 shows a block diagram of the disparity compensated view filtering wavelet based multiview image encoder. From Fig. 3, it can be seen that five gray scale images, representing five views of a scene, are input to the multiview image encoder. The encoder first applies a two level 5/3 lifting based wavelet transform between the input images. It then performs one level Haar lifting based wavelet transform between the resulting basebands. Figure 4 shows the implementation of the wavelet transform between the input images. After Disparity Compensation Views Filtering (DCVF), most of the images energy is concentrated into the baseband, the LLL-image. As the information in the high-pass subband images, L-, LH- and LLH-images, mostly corresponds to the occluded areas, they carry small portion of the images' energy. The resulting baseband image LLL-image then undergoes three levels of two dimensional 5/3 spatial wavelet decomposition, which decomposes the LLL-

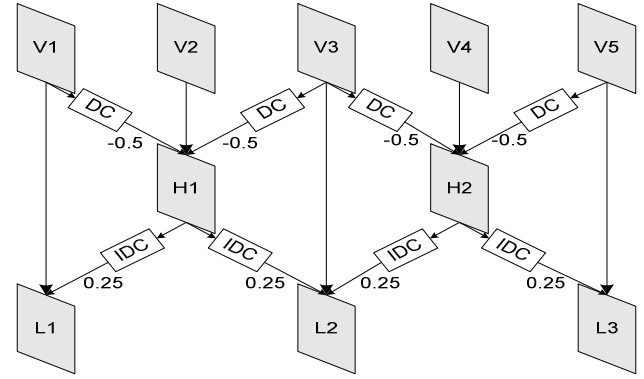


Fig. 1. Implementation steps of the 5/3 lifting based wavelet transform.

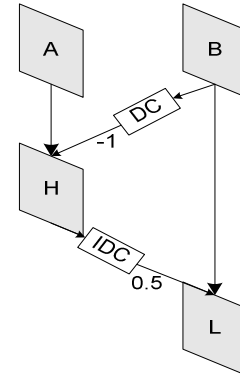


Fig. 2. Implementation steps of the Haar lifting based wavelet transform.

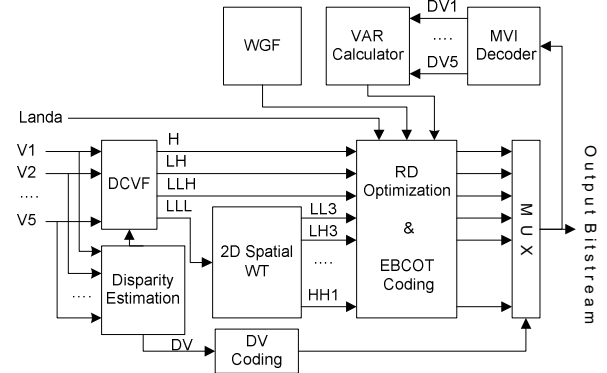


Fig. 3. Disparity compensated view filtering wavelet based multiview image encoder.

image into its ten subbands. The resulting spatial subbands and high-pass images are then considered for optimal bit allocation before being coded. The ultimate goal of the bit allocation among the subbands in image compression schemes is to assign a certain number of bits R_i from the total bit budget R to each of the subbands in such a way that minimize the global distortions between the original and the reconstructed image data.

In this paper, the optimal bit allocation technique presented in [5] is used to assign bits among the different subbands in

order to minimize the overall distortions of the reconstructed images and also to balance the distortion among the reconstructed images in such a way to minimize the variance of the PSNR of the reconstructed images. Therefore the transform gain for each subband is first calculated. The transform gains are then used along with the Lagrangian optimization to perform optimal bit allocation among the subbands. Since the transform gain depends on the actual wavelet decomposition structure, the transform gain for different subbands based on the wavelet decomposition structure of the proposed codec are calculated and listed in Tables I and II. The 5/3 filter, i.e. (1/2, 1, 1/2) for the low-pass synthesis and (-1/8, -1/4, 3/4, -1/4, -1/8) for the high-pass synthesis, and Haar filter, i.e. (1, 1) for the low-pass synthesis and (1/2, -1/2) for the high-pass synthesis, are used for the transform gain calculation. Since the spatial decomposition is performed only on the resulting LLL-image, the transform gain of the resulting spatial subbands can be calculated by direct multiplication of the spatial wavelet gains and the transform gain of the LLL-image. The Lagrangian distortion optimization algorithm for the proposed codec is as follows: Let's assume $D_i(R_i)$ be the distortion of the subband i at bitrate R_i . Hence, the bit allocation among the subbands can be written as:

$$D(R) = \sum_{i=0}^N w_i D_i(R_i) \quad \text{under the restriction of} \quad \sum_{i=0}^M R_i \leq R \quad (2)$$

where R is the bit budget and w_i is the transform gain of subband i . With Lagrangian optimization, this is equal to the minimization of the cost function in Equation 3.

$$J(R_i) = D_i(R_i) + \frac{\lambda}{w_i} R_i \quad \forall i \in \{0, \dots, N\} \quad (3)$$

where λ is the Lagrange multiplier. Minimization of Equation 3 with respect to R_i results in,

$$\frac{\partial D_i(R_i)}{\partial R_i} = -\frac{\lambda}{w_i} \quad \forall i \in \{0, \dots, N\} \quad (4)$$

Therefore points with equal slope on the weighted RD-curves are defined the optimal rate for different subbands. The EBCOT software was used to generate the rate distortion curves and compress the subbands at optimal bitrates. Finally a multiplexer put the resulting bitstream of the different subbands together with the coded DVs and generates the output bitstream.

During the implementation of the proposed codec it was found that the current form of the bit allocation among the subbands can result in having different image quality among the reconstructed images of a set of multiview images, which is not acceptable in many applications. This is due to the fact that the Lagrangian RD optimization algorithm minimizes the global distortion and does not balance the distortion among the different views. To reduce the PSNR variance of the reconstructed images, a weighting factor is introduced to

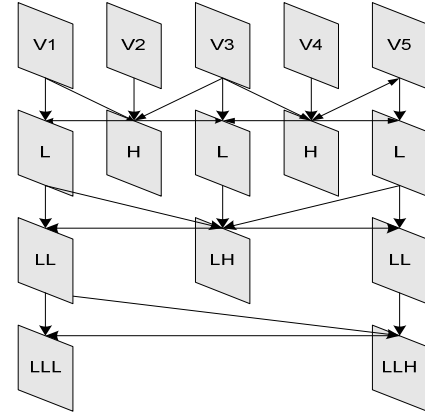


Fig. 4. Implementation of wavelet transform between multiview images in the proposed codec.

TABLE I
TRANSFORM GAIN OF THE DCVF FOR GLOBAL RATE DISTORTION OPTIMIZATION.

WG-H	WG-LH	WG-LLH	WG-LLL
0.7188	0.9219	1.0625	6.7500

TABLE II
TRANSFORM GAIN OF THE SPATIAL DECOMPOSITION FOR GLOBAL RATE DISTORTION OPTIMIZATION.

Subband	WG-LL	WG-LH	WG-HL	WG-HH
Level 1	2.2500	1.0781	1.0781	0.5166
Level 2	7.5625	2.5352	2.5352	0.8499
Level 3	28.8906	8.5244	8.5244	2.5152

generate different λ s for the high-pass images and spatial subbands as shown in Equation 5:

$$\lambda_{view} = k \cdot \lambda_{Spatial}$$

$$\frac{\partial D_i(R_i)}{\partial R_i} = -\frac{\lambda_{Spatial}}{w_i} \quad \forall i \in \{0, \dots, N\} \quad i \text{ addresses a spatial subband}$$

$$\frac{\partial D_j(R_j)}{\partial R_j} = -\frac{\lambda_{view}}{w_j} \quad \forall j \in \{0, \dots, M\} \quad j \text{ addresses a view subband}$$

(5)

and then a steepest descent algorithm is used to adjust the value of k in order to minimize the PSNR variance of the reconstructed images.

4. EXPERIMENTAL RESULTS

In order to evaluate the performance of the proposed codec, frames number 53, 70 and 32 from different views of the Damien, Lydia and Golasa multiview test sequences were taken, respectively, and three sets of multiview test images were generated [6]. Two sets of experiments were then performed on the selected views: (i) the basic architecture of

the proposed codec was used to code the three sets of multiview images using the rate distortion optimization algorithm presented in Equation 5 at $\lambda_{Spatial}$ s of 0.1, 1 and 10 to 100 with steps of 10 and $\lambda_{Spatial} / \lambda_{View}$ s of 0.1 to 2 with steps of 0.1. The variance of the decoded images for each set of the $\lambda_{Spatial}$ and λ_{View} was then calculated. Figure 5 shows a 3D representation of the resulting PSNR variance for the Damien images. From Figure 6 it is clear that the minimum PSNR variances of the decoded images are achieved at different combinations of $\lambda_{Spatial}$ and λ_{View} rather than the general combination, $\lambda_{Spatial} = \lambda_{View}$. This experiment was also repeated on the two other sets of multiview images. Their results confirmed the results presented for the Damien images in this paper, (ii) the proposed codec, which uses the steepest descent algorithm to minimize the PSNR variance of the decoded images, and the basic structure of the codec which uses a general form of the RD optimization were then used to compress the three sets of multiview images. Figure 6 shows the resulting PSNR variance of the two codecs for Damien, Lydia and Golasa multiview images. From Figure 6, it is obvious that the proposed codec always gives lower PSNR variance at all λ s, where there is a direct relationship between the λ s and compression ratios. The coding performance of the proposed codec was not discussed in this paper as it has been shown in several papers that this type of codecs outperform the JPEG2000 codec [1].

5. CONCLUSION

In this paper a disparity compensated view filtering wavelet based multiview image codec was presented. The proposed codec uses a lifting based wavelet transform to decompose the multiview images into a number of low and high-frequency images. It then applies a 2D DWT to further decompose the low-frequency image into its subbands. The general form of the RD optimization algorithm was then modified and used along with the steepest descent algorithm to assign bits among the different subbands in such a way that the PSNR variance of the decoded images is minimized. Experimental results have shown that the general form of the RD optimization preassembly does not produce lower PSNR variance among the decoded images at all compression ratios. Results have also shown that the proposed coding scheme gives lower PSNR variance among the decoded images than the general codecs at all compression ratios.

ACKNOWLEDGEMENT

This work was supported by EPSRC under grant number GR/T11883/01.

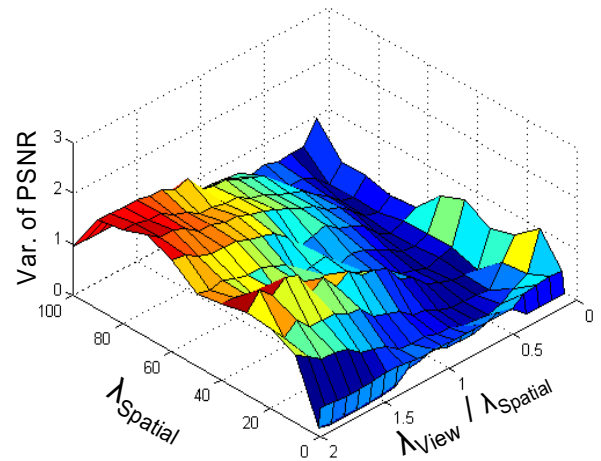


Fig. 5. Variance of the PSNR of the decoded Damien multiview images when using different λ_{View} in coding the H-images.

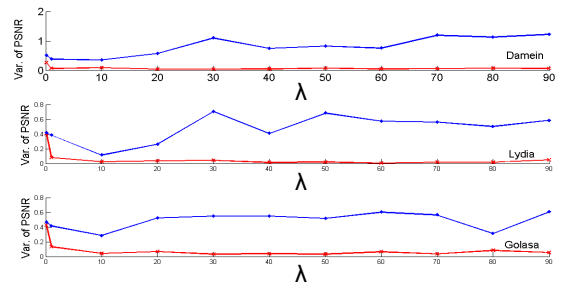


Fig. 6. Variance of the PSNR of the decoded multiview images using a) $\lambda_{Spatial} = \lambda_{View}$ (blue curves) and b) the proposed codec (red curves) for Damien, Lydia and Golasa multiview images at different $\lambda_{Spatial}$ from top to bottom, respectively.

REFERENCES

- [1] N. Anantrasirichai, C. N. Canagarajah and D. R. Bull, "Multi-View Image Coding with Wavelet Lifting and In-Band Disparity Compensation", *ICIP*, Geneva, September 2005.
- [2] A. Sheikh Akbari, N. Canagarajah, D. Redmill and D. Bull, "Multiresolution, Adaptive Vector Quantization and Perceptual Based Multiview Image Codec", *The 26th Picture Coding Symposium (PCS)*, Lisbon, Portugal, November 2007.
- [3] A. Secker and D. Taubman, "Lifting-Based Invertible Motion Adaptive Transform (LIMAT) Framework for Highly Scalable Video Compression", *IEEE Transaction on Image Processing*, vol. 12, no. 12, December 2003.
- [4] J.W. Garbas, U. Fecker, T. Tobias and A. Kaup, "4D Scalable multi-View Video Coding Using Disparity Compensated View Filtering and Motion Compensated Temporal Filtering", *International Workshop on Multimedia Signal Processing (MMSP2006)*, Canada, October 2006.
- [5] B. Usevitch, "Optimal bit allocation for biorthogonal wavelet coding", *Proceeding Data Compression Conference*, March 1996.
- [6] A. Sheikh Akbari, N. Canagarajah, D. Redmill, D. Bull and D. Agrafiotis, "A novel H.264/AVC based multiview video coding scheme", *3DTV conference*, Kos Island, Greece, May 2007.